

Attorney Docket No. 5577-230  
Application Serial No. 09/833,925  
Filed: April 12, 2001  
Page 10

## REMARKS

Applicants appreciate the Office Action of September 17, 2004. Applicants respectfully submit that the pending claims are patentable over the cited references for at least the reasons discussed herein. Accordingly, allowance of the pending claims is respectfully requested in due course.

### The Section 102 Rejections

#### A. The Independent Claims are Patentable over the Cited Reference

Claims 1-6, 10, 11, 16-19, 23-25 and 30-33 stand rejected under 35 U.S.C. 102(b) as being anticipated by *A Robust and Scalable Internet Server* by Dahlin *et al.* (hereinafter "Dahlin"). See Office Action, page 2. Applicants respectfully submit that many of the recitations of these claims are neither disclosed nor suggested by the Dahlin for at least the reasons discussed herein.

Dahlin discusses the Eddie approach. In the EDDIE approach a distributed application called "IP Address Migration Application" controls all IP addresses in the cluster. See Dahlin, page 2, column 1. The cluster is connected via a shared-medium LAN. See Dahlin, page 2, column 1. IP address aliasing is used to provide addresses to individual applications over a single adapter, and these aliases are located via Address Resolution Protocol (ARP) and ARP caches in the TCP/IPs. See Dahlin, page 6. The application monitors all server applications and hardware, and reallocates aliased IP addresses in the event of failure to surviving adapters and nodes. This approach allows applications of a failing node to be distributed among surviving nodes. See Dahlin, page 6

In contrast, Claim 1 recites:

A method of distributing workload between a plurality of servers, the method comprising:

receiving a plurality of requests over a first connection;

parsing the plurality of requests to determine application layer information associated with each of the plurality of requests;

selecting destination servers for corresponding ones of the plurality of requests based on the determined application layer information associated with each of the plurality of requests; and

distributing the plurality of requests to the corresponding selected destination servers over a plurality of second connections associated with respective ones of the

Attorney Docket No. 5577-230  
Application Serial No. 09/833,925  
Filed: April 12, 2001  
Page 11

destination servers.

Claims 30 and 32 contain corresponding system and computer program product recitations, respectively. Applicants respectfully submit that at least the highlighted recitations of, for example, Claim 1 are neither disclosed nor suggested by the Dahlin.

The Office Action points to the following portion of Dahlin as teaching the recitations of Claim 1, among others:

A server built with the Generic Internet Server Framework is divided into two parts; a front-end and a back-end. Front-ends typically listen to several UDP and/or TCP sockets and parses any incoming data. Front-ends analyze incoming requests and schedule them to be run by back-ends either on the same host or on other hosts. A front-end could even decide to schedule requests to back-ends on remote LANs if this is beneficial, e.g. due to extreme overload on all local back-ends or because of services only being available on back-ends on a remote LANs.

See Dahlin, page 4, column 1, fourth full paragraph (emphasis added). In other words, the cited portion of Dahlin discusses a front-end that receives data through the UDP and/or TCP sockets and parses the received data. The parsed data received through the different sockets may then be provided to one or more back-ends. The Office Action implies that since the cited portion of Dahlin discusses "listening to several UDP and/or TCP sockets," Dahlin teaches receiving a plurality of requests over a first connection. See Office Action, page 2. Just because Dahlin listens to several UDP and/or TCP sockets does not mean that more than one request will be received over any single connection. Nothing in the cited portion of Dahlin discloses or suggests receiving a plurality of requests over a first connection as recited in Claim 1. In particular, as discussed in the specification of the present invention, the first connection is established and does not close after the first request is received over the connection, thus, many requests may be received over the same first connection. See Specification, page 6, lines 24-29. Nothing in the cited portion of Dahlin discloses or suggests establishing a first connection as recited in Claim 1.

Furthermore, the Office Action implies since the cited portion of Dahlin states that the front-end of Dahlin analyzes the request, Dahlin teaches parsing the plurality of requests (received over the established first connection) to determine application layer information associated with each of the plurality of requests. See Office Action, page 2. Applicants respectfully disagree. Nothing in the cited portion of Dahlin appears to disclose or suggest

Attorney Docket No. 5577-230  
Application Serial No. 09/833,925  
Filed: April 12, 2001  
Page 12

parsing the requests for application layer information as recited in Claim 1.

The Office Action further implies that the since the cited portion of Dahlin discusses scheduling requests to the back-ends, Dahlin teaches selecting destination servers for corresponding ones of the plurality of requests based on the determined application layer information associated with each of the plurality of requests as recited in Claim 1. See Office Action, page 2. Applicants respectfully disagree. As discussed above, nothing in the cited portion of Dahlin discloses or suggests parsing the requests for application layer information, thus, it follows that nothing in the cited portion of the Dahlin discloses using the application layer information to select destination servers. Requests can be scheduled without using application layer information to schedule the requests. Nothing in the cited portion of Dahlin discloses or suggests receiving a plurality of requests over a single connection and distributing the request over a plurality of second connections to the destination servers as further recited in Claim 1. Accordingly, Claims 1, 30 and 32 and the claims that depend therefrom are patentable over Dahlin for at least the reasons discussed herein.

By way of further example, Claim 16 recites:

A method of distributing workload between a plurality of servers, wherein each of the plurality of servers is executing an instance of an application which communicates over a network such that each of a plurality of HTTP requests within a single HTTP 1.1 connection to the application may be distributed to any one of the plurality of servers, the method comprising:

defining a subset of the plurality of servers which are to receive HTTP requests having an indication of high priority;

establishing an HTTP 1.1 connection responsive to receiving a request for an HTTP 1.1 connection to the application over the network;

receiving a first Hypertext Transport Protocol(HTTP) request within the HTTP 1.1 connection;

parsing the first HTTP request to determine if the first HTTP request has an indication of high priority based on application layer information included in the first HTTP request; and

distributing the first HTTP request to one of the subset of the plurality of servers over a first connection if the first HTTP request has an indication of high priority.

Claims 31 and 33 contain corresponding system and computer program product recitations, respectively. Applicants respectfully submit that at least the highlighted recitations are neither

Attorney Docket No. 5577-230  
Application Serial No. 09/833,925  
Filed: April 12, 2001  
Page 13

disclosed nor suggest by the Dahlin.

The Office Action points to the following portion of Dahlin as teaching "each of the plurality of servers is executing an instance of an application" (See Office Action page 5) as recited in Claim 16:

Eddie contains of a number of unique components which together makes it possible to build extremely robust and scalable Internet servers. Each component can be used separately but component cooperation gives the server even better characteristics.

See Dahlin, column 2, paragraph 1. Nothing in the cited portion of Dahlin discloses or suggests a plurality of servers each executing an instance of an application. The Office Action implies that since component 2 of Dahlin "routes the different parts of the request to the hosts most suited to answer them" (See Dahlin, column 2), Dahlin teaches defining a subset of the plurality of servers as recited in Claim 16. Applicants respectfully disagree. The fact that Dahlin routes different parts of the requests to different hosts, does not disclose or suggest defining a subset of servers as recited in Claim 16. Nothing in the cited portion of Dahlin discloses or suggests defining a subset of the plurality of servers which are to receive HTTP requests having an indication of high priority as recited in Claim 16.

The Office Action points to the same portion of Dahlin as teaching parsing the first HTTP request to determine if the first HTTP request has an indication of high priority based on application layer information included in the first HTTP request as recited in Claim 16. Again, the fact that Dahlin routes different parts of the requests to different hosts, does not disclose or suggest parsing the request to determine if the request has an indication of high priority based on application layer information. Finally, as Dahlin does not disclose or suggest parsing the request to determine if the request has an indication of high priority, it follows that nothing in the cited portion of Dahlin discloses or suggests distributing the first HTTP request if the first HTTP request has an indication of high priority as further recited in Claim 16. Accordingly, Claims 16, 31 and 33 and the claims that depend therefrom are patentable over Dahlin for at least the reasons discussed herein.

B. Many of the Dependent Claims are Separately Patentable

As discussed above, the dependent claims are patentable at least per the patentability of

Attorney Docket No. 5577-230  
Application Serial No. 09/833,925  
Filed: April 12, 2001  
Page 14

the independent claims from which they depend. Many of the dependent claims are also separately patentable.

For example, Claim 3 recites:

A method according to Claim 1, wherein parsing the plurality of requests comprises:  
determining a start point and an end point for each of the plurality of requests within the first connection; and  
identifying application layer information within each of the plurality of requests.

Claim 23 contains similar method recitations. The Office Action states that the following portion of Dahlin teaches the recitations of Claim 3:

A server built with the Generic Internet Server Framework is divided into two parts; a front-end and a back-end. Front-ends typically listen to several UDP and/or TCP sockets and parses any incoming data. Front-ends analyze incoming requests and schedule them to be run by back-ends either on the same host or on other hosts. A front-end could even decide to schedule requests to back-ends on remote LANs if this is beneficial, e.g. due to extreme overload on all local back-ends or because of service only being available on back-ends on a remote LANs.

See Dahlin, page 4, column 1, fourth full paragraph (emphasis added). Nothing in the cited portion of Dahlin discloses or suggests the recitations of Claim 3. Accordingly, Claims 3 and 23 are separately patentable over Dahlin for at least these additional reasons.

By way of further example, Claim 10 recites:

A method according to Claim 1, wherein distributing the plurality of requests comprises:  
determining if a second connection associated with a selected destination servers exists;  
establishing the second connection to the selected destination server if the second connection does not exist;  
distributing a request to the selected destination servers over the second connection; and  
repeating the determining, establishing and distributing for each of the plurality of requests.

Claim 24 contains similar recitations. Applicants respectfully submit that at least the highlighted portions of Claim 10 are neither disclosed nor suggested by the cited portion of Dahlin. In particular, the Office Action points to the following portion of Dahlin at teaching the highlighted

Attorney Docket No. 5577-230  
Application Serial No. 09/833,925  
Filed: April 12, 2001  
Page 15

recitations of Claim 10 (*See Office Action, page 4*):

The first time a client accesses a DNS server, the DNS server will not have response time data. It then returns the IP server or list of servers with the lowest load in a WAN.

*See Dahlin, page 1, column 2.* Nothing in the cited portion of Dahlin discloses or suggests determining if a second connection exists and establishing the second connection if the second connection does not exist. Accordingly, Applicants respectfully submit for at least these additional reasons that Claims 10 and 24 are separately patentable over Dahlin.

By way of final example, Claim 17 recites:

A method according to Claim 16, further comprising:  
distributing the first HTTP request to a server other than a server in the subset of the destination servers if the first HTTP request does not have an indication of high priority.

The Office Action points to the recitations of Dahlin at page 1, column 2 as teaching the recitations of Claim 17. Nothing in the cited portion of Dahlin discloses or suggests determining the recitations of Claim 17. Accordingly, Applicants respectfully submit that Claim 17 is separately patentable over Dahlin for at least these additional reasons.

#### The Section 103 Rejections

Claims 7-9, 12-15, 20-22 and 26-29 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Dahlin in view of *Efficient Support for P-HTTP in Cluster-Based Web Servers* by Mohit Aron *et al.* (hereinafter "Aron"). As discussed above, the dependent claims are patentable over the cited references at least per the patentability of the independent base claims from which they depend. However, many of the dependent claims are also separately patentable for at least the reasons stated herein.

In particular, the Office Action admits that Dahlin fails to disclose or suggest all of the recitations of Claims 7, 12-15 and 26-29. *See Office Action, pages 9-13.* However, the Office Action points to Aron as providing the missing teachings. For example, Claim 7 recites:

A method according to Claim 1, wherein selecting destination servers for corresponding ones of the plurality of requests comprises:  
determining if the determined application layer information associated with

Attorney Docket No. 5577-230  
Application Serial No. 09/833,925  
Filed: April 12, 2001  
Page 16

each of the plurality of requests is relevant application layer information;  
selecting one of a subset of the destination servers if the application layer  
information associated with each of the plurality of requests is relevant application  
layer information; and  
selecting a destination server other than a destination server in the subset of  
the destination servers if the application layer information associated with each of  
the plurality of requests is not relevant application layer information.

The Office Action states that the fact that component 2 of Dahlin "routes the different parts of the request to the hosts most suited to answer them" (See Dahlin, page 1, column 2) teaches the determining step of Claim 7. See Office Action, page 9. Applicants respectfully submit that nothing in the cited portion of Dahlin discloses or suggests determining if the determined application layer information associated with each of the plurality of requests is relevant application layer information as recited in Claim 7. In fact, as discussed above, nothing in the cited portions of Dahlin specifically states that application layer information is ever determined. Furthermore, the Office Action admits that Dahlin does not teach the selecting steps of Claim 7. Accordingly, nothing in the cited portion of Dahlin discloses or suggests the determining or the selecting steps of Claim 7.

The Office Action points to Figure 1 of Aron as providing the missing teachings. The text corresponding to Figure 1 of Aron recites:

Figure 1 illustrates the principle of LARD in a cluster with two back-ends and a working set of three targets (A, B and C) in the incoming request stream. The front-end directs all requests for A to back-end 1, and all requests for B and C to back-end 2. By doing so, there is an increased likelihood that the request finds the requested target in the cache at the back-end.

Aron, page 2, column 2. In other words, Figure 1 of Aron illustrates all requests for A being routed to a first back-end and all requests for B and C being routed to a second back-end. In contrast, Claim 7 recites determining if the determined application layer information is relevant application layer information, selecting one of a subset of the destination servers if the application layer information associated with each of the plurality of requests is relevant application layer information and selecting a destination server other than a destination server in the subset of the destination servers if the application layer information associated with each of the plurality of requests is not relevant application layer information. Nothing in

Attorney Docket No. 5577-230  
Application Serial No. 09/833,925  
Filed: April 12, 2001  
Page 17

the cited portion of Aron discloses or suggests these recitations of Claim 7. In fact, the destination of the requests in Aron is predetermined and is not based on the content of the request, *i.e.* not based on the presence or lack of presence of relevant application layer information.

Accordingly, neither Dahlin nor Aron either alone or in combination disclose or suggest many of the recitations of Claim 7 set out above. Furthermore, there is no motivation or suggestion to combine the cited references as suggested in the Office Action. As affirmed by the Court of Appeals for the Federal Circuit in *In re Sang-su Lee*, a factual question of motivation is material to patentability, and cannot be resolved on subjective belief and unknown authority. See *In re Sang-su Lee*, 277 F.3d 1338 (Fed. Cir. 2002). It is improper, in determining whether a person of ordinary skill would have been led to this combination of references, simply to "[use] that which the inventor taught against its teacher." *W.L. Gore v. Garlock, Inc.*, 721 F.2d 1540, 1553, 220 U.S.P.Q. 303, 312-13 (Fed. Cir. 1983).

The Office Action states:

It would have been obvious to a person of ordinary skill in the art at the time of the applicant's invention to modify the teaching of Dahlin to include "selecting one of a subset of the destination servers if the application layer information associated with each of the plurality of requests is relevant application layer information; and selecting a destination server other than a destination server in the subset of destination servers if the application layer information associated with each of the plurality of requests is not relevant application layer information" for the requests to find targets.

See Office Action, page 10. This motivation is a motivation based on "subjective belief and unknown authority", the type of motivation that was rejected by the Federal Circuit in *In re Sang-su Lee*. In other words, the Office Action does not point to any specific portion of the cited references that would induce one of skill in the art to combine the cited references as suggested in the Office Action. If the motivation provided in the Office Action is adequate to sustain the Office's burden of motivation, then anything that would allow the "requests to find targets" would render a combination obvious. This cannot be the case. Accordingly, the statement in the Office Action with respect to motivation does not adequately address the issue of motivation to combine as discussed in *In re Sang-su Lee*. Thus, it appears that the Office Action gains its alleged impetus or suggestion to combine the cited references by hindsight reasoning informed

Attorney Docket No. 5577-230  
Application Serial No. 09/833,925  
Filed: April 12, 2001  
Page 18

by Applicants' disclosure, which, as noted above, is an inappropriate basis for combining references.

Furthermore, as discussed above, Dahlin discusses the Eddie approach, which allows applications of a failing node to be distributed among surviving nodes. Aron, on the other hand, discusses efficient support for P-HTTP in cluster based web servers as recited in the title. Nothing in the cited references would motivate a person of skill in the art to combine Dahlin and Aron. Furthermore, even if Dahlin and Aron could be properly combined, the combination of Dahlin and Aron would not result in the recitations of the pending claims for at least the reasons discussed above.

The Office Action provides similar motivations with respect to Claims 12-15 and 26-29. See Office Action, pages 9-13. Thus, Applicants submit that Dahlin and Aron cannot be properly combined to teach the recitations of Claim 12-15 and 26-29 for at least the reasons discussed above with respect to Claim 7. Accordingly, Applicants respectfully submit that 7-9, 12-15, 20-22 and 26-29 are separately patentable over the cited combination for at least these additional reasons.

### CONCLUSION

Applicants respectfully submit that pending claims are in condition for allowance for at least the reasons discussed above. Thus, allowance of the pending claims is respectfully requested in due course. Favorable examination and allowance of the present application is respectfully requested.

Respectfully submitted,



Elizabeth A. Stanek  
Registration No. 48,568

USPTO Customer No. 46589  
Myers Bigel Sibley & Sajovec, P.A.  
Post Office Box 37428  
Raleigh, North Carolina 27627  
Telephone: 919/854-1400  
Facsimile: 919/854-1401